Using timbre to predict musical genre: Promising solution or dead end?

Cory McKay
McGill University
Montréal, Canada

Central question

- How useful is timbre in automatically classifying music?
  - Useful by itself?
  - Useful in combination with other information?
  - Not useful at all?
- Genre classification used as a case study

Presentation overview

- Overview of automatic genre classification
- The jMIR toolkit
- Experiment 1:
  - Combining features extracted from audio, symbolic and cultural data
- Experiment 2:
  - Focusing on features extracted from symbolic data
- Final comments

What is genre classification?

- Using computers to automatically associate music with genre class labels
- Genre labels can be broad:
  - Jazz, classical, rock, rap, etc.
- Genre labels can be narrow
  - Microsound, chiptunes, glitch, IDM, etc.

Why is genre classification useful?

- Music consumers still browse music by genre (Lee and Downie 2004)
  - Consumers can be very disobedient to the wishes of some MIR researchers
- Genre can provide important musicological and music theoretical insights into how humans organize and classify music at a high level
  - Fabbri, Frith, Brackett, etc.
- Genre classification shares characteristics with other types of music classification
  - Mood, listening scenario, performer, composer, etc.
  - An interestingly hard problem whose solution may provide wide-ranging insights into other classification problems

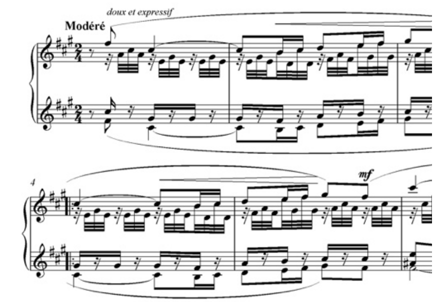
How is gemre classification done?

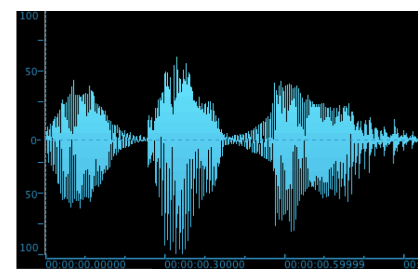
- Collect labeled ground truth training and testing data
  - Possibly involving structured class ontologies
- Extract features from this data
- Build a classification model using supervised learning algorithms
- Validate the model

- Similar methodology to many other kinds of automatic music classification

Main feature sources

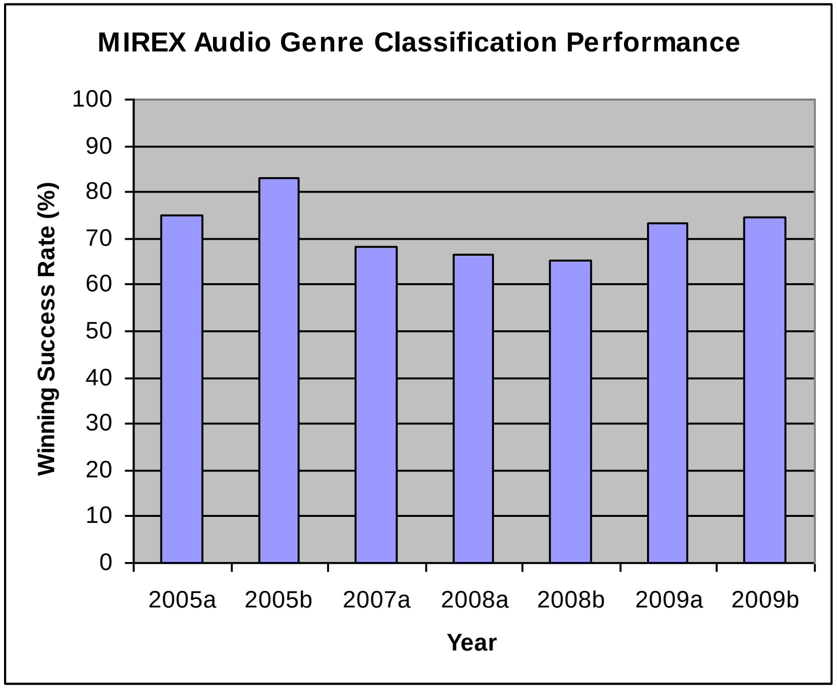
- Symbolic recordings
  - e.g. MIDI or Humdrum files
- Cultural data
  - e.g. web text or metadata tags
- Audio recordings
  - e.g. MP3 or .wav files
  - Traditional source of timbral features
- Others: lyrics and album art





How well can we do?

- The MIREX contest is the best way to compare performance
- Best results to date:
  - 6 classes: 82.9% (2005b)
  - 10 classes: 75.1% (2005a)
- Differences between datasets make it different to fairly compare results, but:
  - There is no evidence of significant improvement from year to year



Note: 2005b involved 6 genres and all other runs involved 10 genres

Commonalities in approaches?

- Relatively easy datasets
  - Genre classes tend to be quite different from one another
  - 10 genre classes are not very many
- Some diversity in machine learning strategies
  - Including some very interesting and effective approaches (and some less so)
- Features associated primarily with timbre…
  - Although some simple features associated with pitch and rhythm are used as well

"Uh oh" says timbre

- Are timbral features the limiting factor?

- Let's look at some experimental data…

Commercial interlude: jMIR

Software tools used: jMIR

- jMIR is a free and open-source Java software suite designed for general music classification research:
  - jAudio: Audio feature extraction
    - 26 core features + metafeatures and aggregators
  - jSymbolic: Feature extraction from MIDI files
    - 111 mostly original features
  - jWebMiner: Cultural feature extraction
    - Uses search engine co-occurrence page counts
  - ACE: Meta-learning classification system
    - 7 machine learning and 3 dimensionality reduction algorithms

More on jMIR

- jMIR also includes other components
  - ACE XML
  - Codaich
  - jMusicMetamanager
  - jMIRUtilities
  - Bodhidharma MIDI
- More information:
  - jMIR's components have each been described individually in various publications
  - jmir.sourceforge.net
  - cory.mckay@mail.mcgill.ca

We now return to our feature presentation

Experiment 1(ISMIR 2008)

- Can combining features extracted from audio, symbolic and/or cultural sources significantly improve automatic music classification performance?
  - Intuitively, they each seem to contain very different kinds of information
- Can this help us break the seeming genre classification performance ceiling?

Experimental methodology

- Extracted features from separate audio, symbolic and cultural sources of data
  - Corresponding to the same musical pieces
- Compared genre classification performance of each of the 7 possible subsets of these 3 feature groups
  - Audio, Symbolic + Audio, Cultural, Symbolic + Cultural + etc.
  - 10-fold cross-validation

Musical dataset used: SAC

- The SAC Dataset was assembled for this experiment
  - Symbolic Audio Cultural
  - 250 recordings belonging to 10 genres
  - Audio and MIDI versions of each recording
    - Acquired separately
  - Accompanying metadata that could be used to extract cultural features from the web

Genres in SAC

- SAC's 10 genres can be collapsed into 5 genres in order to separately evaluate performance on both moderate and small genre taxonomies
  - Facilitates evaluation of misclassification severity

- **Blues:** Modern Blues and Traditional Blues
- **Classical:** Baroque and Romantic
- **Jazz:** Bop and Swing
- **Rap:** Hardcore Rap and Pop Rap
- **Rock:** Alternative Rock and Metal

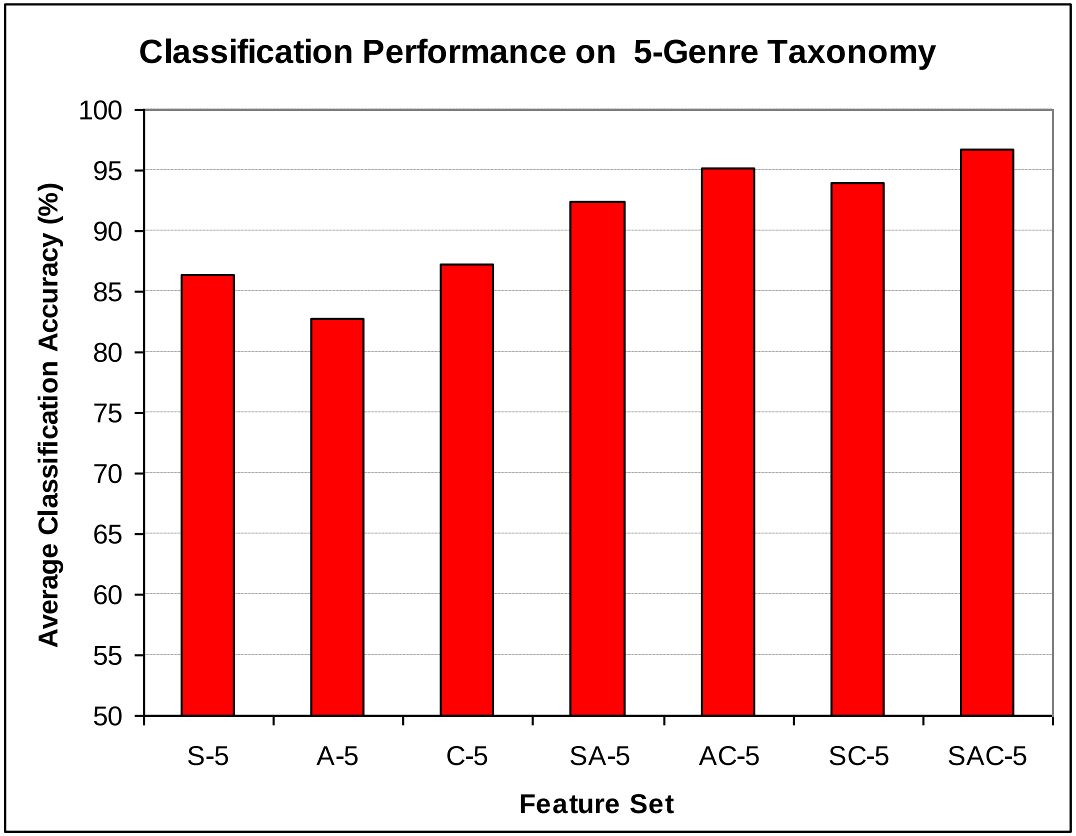
Difficulty of SAC

- Performances of the same song in different genres
- Performances by the same artists in different genres
- 10-genre taxonomy includes pairs of relatively similar genres

- These factors make SAC harder than the typical MIREX datasets
  - More realistic, although still easier than real-world application would require

Results: 5-genre taxonomy

- 3 feature types vs. 1 type
  - 11.3% better
    - A 78% decrease in the error rate
  - Statistically significant
- 3 feature types vs. 2 types
  - 2.3% better
  - Not statistically significant

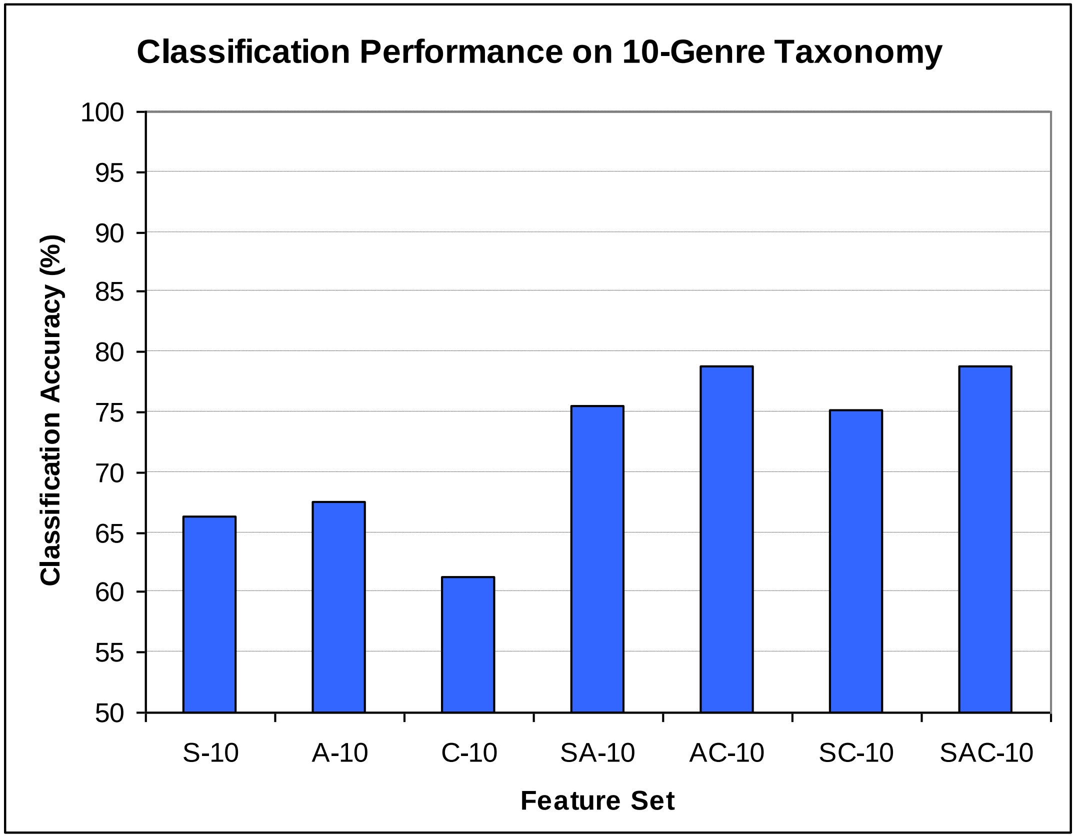


"Uh oh" says timbre, again

- Audio was the worst performing single data type
  - Most (but not all) features extracted from it were timbral

Results: 10-genre taxonomy

- Trends similar to 5-genre results
- 3 feature types vs. 1
  - 13.7% better
    - A 39.3% decrease in the error rate
  - Statistically significant
- 3 feature types vs. 2
  - 2.7% better
  - Not statistically significant



"Yay!" says timbre

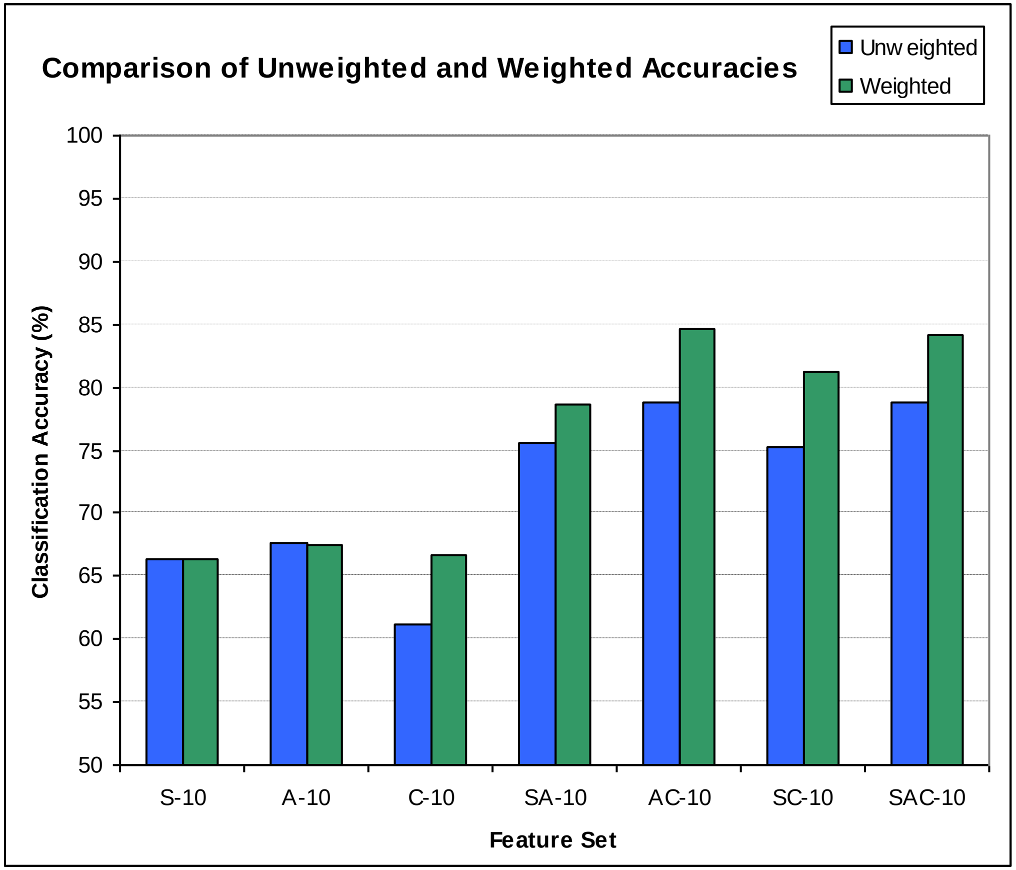
- Audio was the best performing single data type
- Perhaps timbre-based features are not a bridge to nowhere?

Misclassification seriousness

- Misclassification to a similar genre can be less serious than misclassification to a dissimilar genre
  - e.g., John Lennon → Beatles vs. John Lennon → Rihanna
- To investigate this, we calculated weighted classification accuracies for the 10-genre experiments
  - Misclassification within a SAC genre pair: 0.5 error
  - Misclassification outside a SAC genre pair: 1.5 error
- Recall SAC genre pairs:
  - **Blues:** Modern Blues and Traditional Blues
  - **Classical:** Baroque and Romantic
  - **Jazz:** Bop and Swing
  - **Rap:** Hardcore Rap and Pop Rap
  - **Rock:** Alternative Rock and Metal

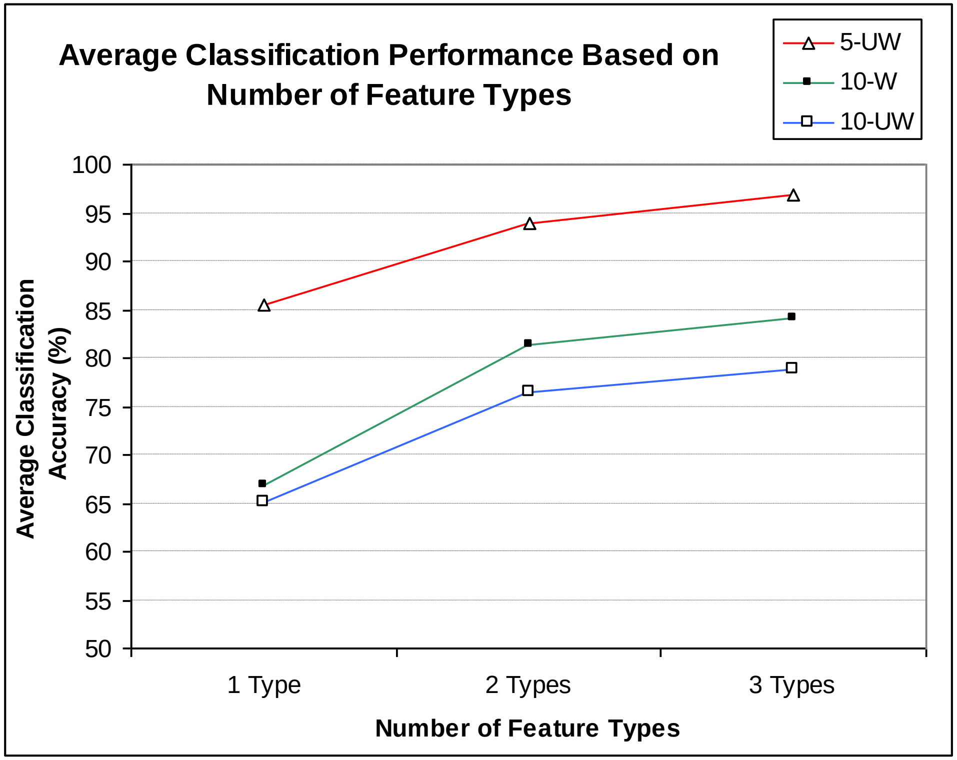
Results: weighted vs. unweighted

- Audio and symbolic
  - No significant difference
  - Although weighted 3% greater than corresponding unweighted when both combined
- Feature groups including cultural features had fewer serious misclassifications than those without cultural features
  - Weighted greater than corresponding unweighted by average of 5.7%
  - Statistically significant



Experiment 1 conclusions

- Combining two or more feature groups improves performance compared to any single feature group
- Using cultural features causes those misclassifications that do occur to be less serious
- The performance ceiling on genre classification performance may not be as low as some have worried



But what about timbre?

- It looks like timbre-based features can play a role, but may be limited by themselves

Experiment 2 (CIM 05)

- An examination of the relative effectiveness of different high-level features in automatic genre classification
- Focused on features extracted from symbolic data
  - MIDI specifically

Software used

- Used jMIR Bodhidharma
  - The ancestor of jSymbolic and ACE
  - Extracts 111 symbolic features
  - Performs dimensionality reduction using genetic algorithms
    - Binary feature selection
    - Linear feature weighting
  - Learning ensemble utilizes of a combination of flat, hierarchical and round robin strategies
    - Multi-layer perceptrons
    - K-NN

Features

- 111 high-level features implemented:
  - Pitch Statistics
    - e.g. fraction of notes in the bass register
  - Melody
    - e.g. fraction of melodic intervals comprising a tritone
  - Instrumentation
    - e.g. whether modern instruments are present
  - Musical Texture
    - e.g. standard deviation of the average melodic leap of different lines
  - Rhythm
    - e.g. standard deviation of note durations
  - Dynamics
    - e.g. average note to note change in loudness
- 42 more features have been proposed but have not been implemented yet, including features based on chords

Genre ontology

- Performed experiments on two genre taxonomies:
  - Large (38 leaf classes):
    - Tests system under realistic conditions
  - Small (9 leaf classes):
    - For loosely comparing system to other experiments

Large taxonomy

Country
- Bluegrass
- Contemporary
- Trad. Country

Jazz
- *Bop*
  - Bebop
  - Cool
- *Fusion*
  - Bossa Nova
  - Jazz Soul
  - Smooth Jazz
- Ragtime
- Swing

Modern Pop
- Adult Contemp.
- *Dance*
  - Dance Pop
  - Pop Rap
  - Techno
- Smooth Jazz

Rap
- Hardcore Rap
- Pop Rap

Rhythm and Blues
- *Blues*
  - Blues Rock
  - Chicago Blues
  - Country Blues
  - Soul Blues
- Funk
- Jazz Soul
- Rock and Roll
- Soul

Rock
- *Classic Rock*
  - Blues Rock
  - Hard Rock
  - Psychedelic
- *Modern Rock*
  - Alternative Rock
  - Hard Rock
  - Metal
  - Punk

Western Classical
- Baroque
- Classical
- *Early Music*
  - Medieval
  - Renaissance
- Modern Classical
- Romantic

Western Folk
- Bluegrass
- Celtic
- Country Blues
- Flamenco

Worldbeat
- *Latin*
  - Bossa Nova
  - Salsa
  - Tango
- Reggae

Small taxonomy

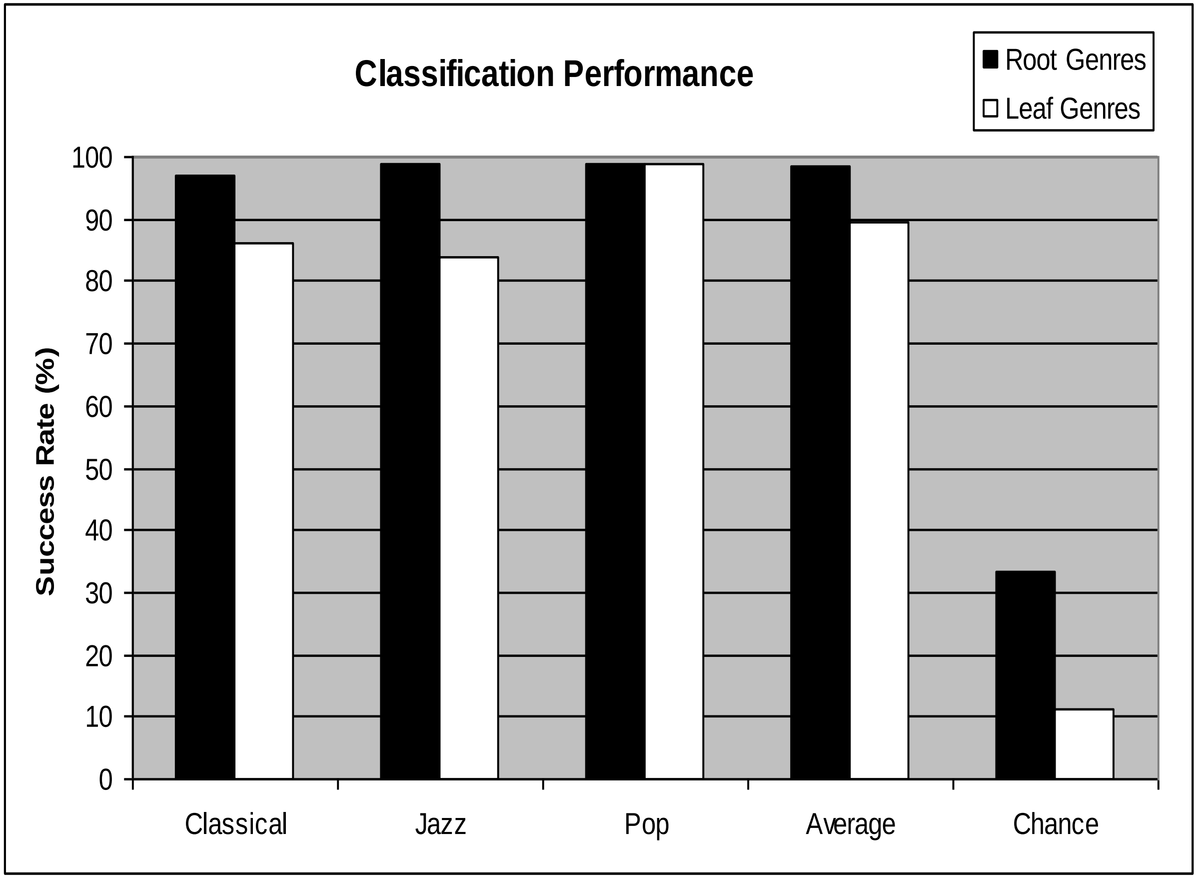
- **Jazz**
  - Bebop
  - Jazz Soul
  - Swing
- **Popular**
  - Rap
  - Punk
  - Country
- **Western Classical**
  - Baroque
  - Modern Classical
  - Romantic

Experimental methodology

- Extracted all features from 950 MIDI files
- Performed GA-based feature weighting
  - Fitness based on classification performance of intermediate trained models
- Classified reserved validation data using the final feature weightings
  - 5-fold cross-validation

Average success rates

- 9 Class Taxonomy
  - Leaf: 90%
  - Root: 98%
- 38 Class Taxonomy
  - Leaf: 57%
  - Root: 81%



Feature performance

Feature Group	Number of Features	Weighting Scaled by Number of Features (%)
Instrumentation	20 (18%)	46.1
Pitch	25 (22%)	24.5
Rhythm	30 (27%)	14.3
Melody	18 (16%)	11.6
Texture	14 (13%)	1.7
Dynamics	4 (4%)	1.6

- Features based on instrumentation were collectively assigned 46.1% of all weightings (after scaling)
  - Even though they only made up 18% of the total features
- At least one instrumentation feature played a major role in almost every classifier in the ensemble
- Two of the top three features were based on instrumentation

Experiment 2 conclusions

- Features based on instrumentation appeared to be very useful
- Caveat:
  - Other features played a dominant role in certain stages of classification
  - The best results were achieved by including a wide variety of features and applying feature selection

But wait… Timbre is great!

- Instrumentation is a high-level abstraction of timbre

Final comments

- Features related to timbre can prove to be very useful in performing automatic music classification
  - At both low and high levels of abstraction
- Timbre-related features seem to be most effective when combined with other kinds of data
- It could be very useful to extract high-level timbral information from audio and use it in high-level features
  - Instrument identification
  - Performance gestures (e.g. bow pressure and speed)
  - Studio audio effects

Acknowledgements

- Funding, past and present:
  - Fonds de recherche sur la société et la culture
  - Social Sciences and Humanities Research Council of Canada
  - The Andrew W. Mellon Foundation
- My co-authours on related papers:
  - Ichiro Fujinaga, John Ashley Burgoyne and Jessica Thompson

- Contact information:
  - cory.mckay@mail.mcgill.ca
  - jmir.sourceforge.net

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

nema

McGill

Schulich School of Music
École de musique Schulich

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

DDMAL